What we learned about making a CD ROM Encyclopedia.

by Kevin Kelly, Whole Earth Review

An electronic Whole Earth Catalog is something we have wanted to do for a long time. CD ROM technology makes it reasonable to imagine doing it now. The keystone for a self-supporting CD ROM Catalog is a practical hypertext tool. This report outlines our investigation of Wildcard in CD ROM as the likely candidate.

We envision a Whole Earth Learning Disc spinning next to a Macintosh. It invites exploration with clear, comprehensible structure. Using Wildcard, the modular unit of expression is a **card** (rather than a page). A card may trigger text, pictures, or sound. In general, there will be one "idea" per card. That is, one paragraph, one picture, one melody, or when video is possible, one scene for each card.

The user digs deeper into the Disc by clicking a path from the general Catalog domains (such as Whole Systems, Communications, Livelihood, and so on) to the specific reviews of particular books, catalogs, products, and software. He sees brief presentation of a useful concept or instrument, and then samples of it, either in pictures or words. In sum, the five or so Wildcards for each item convey the flavor of its character and utility. A further click brings up the mail-ordering information. (A future version could order the item automatically via a telecommunications program).

The Learning Machine

A Wildcard/Whole Earth CD ROM Catalog would be a kinetic, fluid teaching machine. Users become students, learning directly from the content, learning more from the patterns in which the knowledge is connected, and learning most from the teacher behind the machine: themselves. The chief goal of this CD ROM disc is re-learning how to learn. It teaches the meta-skill of how to aquire new skills, burning in with video-game intensity the habit of diving into the unknown. We might call it the Whole Earth Learning Disc.

In the paper editions of the Catalog we are always hampered by limited space. The Learning Disc allows us to enter the extra book excerpts that we rarely had room for in the past. Continued clicking on one review will continue to turn up further details and quotes from the item.

By design, the Whole Earth Catalog rewards exploration immediately. Readers open it up anywhere, find fascination on the spot, and are then lead into unfamiliar areas. This inherent random-access nature, together with a highly evolved layout for compacting information, makes The Whole Earth Catalog an ideal candidate for a medium like CD ROM. The Catalog also utilizes the advantages of ample graphics, a strength that CD ROM feeds on. CD ROM also offers a wonderful way to deal with our handy, but arbitrary classification scheme. Should the map section be found in "Communications," "Whole Systems," or "Land Use"? That 's a question of the past. It will now be found in all three. It will be like looking up a subject by whatever peculiar name you know it as, and finding it there.

Content

The goal is to have the material go as deep as the reader's interest goes, but no further. This disc will not exhaust a reader's interest in each thread, and thus exhaust his limited energy for sitting in front of a computer. In this way, the Learning Disc is different from an encyclopedia, which ordinarily compiles everything that is known into one volume, without selection for interest or excitement. (Encyclopedias do have their uses, of course). In the Learning Disc the emphasis is on1) acquiring practical information, 2) becoming excited about new areas, 3) developing proficiency with the language and tools of real-world endeavors, and 4) mastering the method of learning.

We brew self-education. Most of the people who are working on this project are professional learners. The subjects the Disc would cover in depth are the kind of things we would have liked to come across while we were in college: Where do you find an expert on a topic? What classics should I read with my limited time? How do I write a resume? How do you make a good diagram? What's a good scientific calculator? What are librarians for? How do you research via an on-line database? Which is the best dictionary? Most of the items we

suggest are low-cost, or borrowable. We are speaking to someone on a student budget.

How Big?

Our guess of an appropriate size for a Learning Disc that would be large enough to be interesting, but small enough to be practical, is about 20,000 Wildcards (50 megabytes of storage). That's approximately the size of The Next Whole Earth Catalog.

The bulk of this database already exists on floppy disks in machine readable form -- the typesetting discs from the latest version of the Whole Earth Catalog (EWEC). Ted Kaehler at Apple has translated them into Mac-ready data.

The floppies hold what would be the basis of the Learning Disc -- about 2,000 items, each with examples and access information. As of 1986, these items are the best, most essential tools we know about. Our experience with the currency of this kind of information is that we can expect it to go out of date at the rate of about 20% per year. The addresses and ordering information itself decays at the rate of 5% per year. This means that the EWEC material would need to be updated by 20% to make a new, believable Learning Disc. We estimate that about 15% of the total effort involved in compiling the Catalogs is devoted to researching and maintaining the integrity of its access information -- price, source address, and current availability.

A notable exception to the readiness of the EWEC is that these discs do not contain pictures. The entire realm of illustrations for the Learning Disc would need to be re-inputed, plus the additional new pictures demanded by the 20% new material needed.

Newsworthy

We have plenty of new stock at hand. The chief project at Whole Earth these days is a research blitz aimed at gathering useful recommendations about communication tools and techniques. Called SIGNAL, this "book" takes off where the Whole Earth Software Catalog stopped. Besides evaluating software, SIGNAL will review the upcoming world of desktop copiers, smart phones, backyard

satellite dishes, video camcorders, teleconferencing systems, and international radio broadcasts. In other words, hi-tech, hi-touch communication.

On the other hand, SIGNAL unfashionably promotes basic low-tech communication skills such as writing, public speaking and library research strategies. Without these fundamental crafts how can a student effectively employ hi-speed communication tools? SIGNAL's investigations would bring to the Learning Disc the impact of hands-on learning, making the Disc a wonderful demo of its own message.

How Much Work?

Our exploration of a Wildcard driven Learning Disc began with several brainstorming meetings at Apple, then several weeks of constructing a pilot prototype. Mimicking a regular production run, we built cards from existing editorial material. We designed screen formats, (re-)keyboarded text and digitized pictures. We put together a micro-demo consisting of less than 100 cards to get a measurement of how long CD ROM publishing takes.

Taking into account acclimation time, we estimate that in a large project, the time needed to construct each card is one hour.

One hour per card includes research, data entry, linking, and proofreading. This translates into 20,000 man-hours to erect the full-bore 20,000-card Learning Disc. For comparison, the 416-page EWEC took 12,000 man hours to make. Since it has 1,962 items, this figures out to be 6 hours per item. (One item will have on average 4 or 5 cards). Thus the labour required for creating a database in Wildcard is in line with the time it takes us to generate our printed information.

If we took only the EWEC data captured on the floppies and contributed no up-to-date new material, the estimated time (at one hour per card) would run to 10,000 man-hours.

The fabrication of past Whole Earth Catalogs required 20 full-time people, including the necessary support people. The Learning Disc, and perhaps other CD ROM based projects, would need to divvy up

this work into several new categories of staff. We see the following as vital:

Researcher/Accesser
Text Inputer
Graphic/Sound Person
Designer
Link Manager
Editor

The Link Manager's sole responsibility is to choose, implement and then verify the hypertext links threaded through the Disc. Besides the obvious task of minding the complexity of links and being sure that they function as intended, the Link Manager must exert superhuman restraint to prevent the connections within the database from becoming a tangle of confusing spaghetti. Without someone to force restraint, the unthinking ease of adding links in Wildcard will drag the disc down the slope to "everything is connected to everything" crankiness. It becomes a labyrinth without design.

Resources

The number of people assigned to each task will vary through the project. It is clear that the task would require new styles of work. Presently we share work by passing pieces of paper around by hand in a single large production room. We have not been able to achieve the same efficiency passing floppy discs around. Although we haven't tested an Appletalk net on the project, it seems certain that an arrangement where files can be served electronically to different jobs is the only way to do the Disc on a large scale. A file server would enable about twice as many people to work together, thus halving the time required.

A Disc will require a minimum of six Mac Pluses on location (one per role) taking into account the flexibility of using our existing MS-DOS machines for raw text entry if this became a bind. Other equipment that we project would be needed is:

1 File Server
1 80 megabyte hard disk for server
2 20 megabyte hard disks for graphics and sound
1 Backup tape drive
1 Laserwriter Plus

1 Flatbed scanner
1 Video scanner
1 Frame grabber
1 Preview monitor
1 Dubbing tape deck
1 Sound digitizer
Many yards of Appletalk cable

The Importance of Graphics

Good graphics are paramount for the success of a CD ROM Disc. The few attempts we have seen of complex databases trying to be friendly without pictures have been deadly. Much of the bewitching power of the Whole Earth Catalog stems from its many, abundant illustrations. It would be stupifyingly dull without them.

We have worked out a rule of thumb that successful print media seem to adhere to: keep the pictures in a ratio of one to one to text. For every inch of words, have an inch of diagrams. Magazines like Scientific American that regularly present their information as 50% graphic remain the standard of conceptual clarity. Whole Earth publications are 30-40% graphics. We would consider 30% the minimum for a CD ROM disc: for every two cards of text, there should be one card of pictures. Anything less will probably be suffocating.

The graphic capabilities of the Macintosh are just barely sufficient for a truly absorbing Disc. We tried out several mechanisms for digitizing pictures to port into Wildcard. None were very satisfactory.

The first general category we tried were video inputs. We aimed digitizing video cameras at ordinary photographs and hard objects of the kind that would appear in the Learning Disk. Good results required much fiddling and redoing. Even at best the final images were representational and not information-dense. We rented two professional video cameras to see if their superior optics would help. They didn't.

The second choice was a flat-bed Abaton digitizer. We found this got its best result from high-contrast originals, whereas the video

cameras out-preformed it on gray half-tones. Only about 20% of our graphic stock is the kind of high contrast image this machine is suited for. A further obstacle is the Abaton's requirement for unattached pages that can be fed into the roller device. Almost none of our artwork comes like that, or any other artwork we can imagine using. As a compromise we tried xeroxing the artwork on a copy machine and then digitizing the copy. Although this adds a second reproduction generation, the degradation was not significant compared to the gross abstraction that digitizing creates.

Since our sources for illustrations come in all varieties, it will be necessary to employ both videos and flat-bed digitizers. Switching between the two systems will require increased effort, but until a more flexible graphic input is found, we see no alternative.

If we stick with the instruments available now, the graphic quality of the Disc will be on a lower order than what we print on paper. Wildcard/CD ROM illustrations will have to be emblematic and diagrammatic rather than fine resolution documents. We tried retouching the digitized images but found that 1) it didn't help much, and 2) it took enormous amounts of time as well. Sizing pictures is also quite cumbersome at the moment. We concluded that preparing the illustrations and transposing them onto their final cards will probably devour over half of the project's time.

We envision employing sound throughout the Disc, most notably in the music section, but also in other places, such as the wildlife and languages parts. We did not make a dry-run of music samples, but from our experience in transferring medias, we estimate that it will follow the pattern of taking about one hour per snip to find, record, track, identify, link, and index. In our estimates of the gross time needed for the project we counted sound samples as picture cards.

How Much Will It Cost?

To keep on a self-correcting path, we are aiming to put together a prototype demonstration disc of about 1,000 Wildcards. This will produce a base that is large enough for a impressive demonstration, although it will be far short of being a stand-alone tool. It will present all the fabulous capabilities of hypertext, as well as the captivating abilities to bring forth music and pictures, and it will do it with intriguing, interest-proven material.

To develop a demostration disc of 1,000 cards based on the material of the Essential Whole Earth Catalog would cost about $52,000. This pays for the learning curve, for convention development, for the tools and jigs we need to invent, for updating the accuracy of the content, and for labor. (Apple would supply the hardware). We are figuring 1,000 hours to make the cards (at one hour per card), adding about 50% of that again for start-up woes, and billing $35 per hour, which covers our time and overhead.

The $52,000 will create a ready-to-run CD ROM disc that is fun, designed for users by a customer-oriented publisher, and loaded with significant material that showcases the unique temperament of CD ROM. It would also buy the nitty-gritty data of the experience we'd gain in assembling a Wildcard/CD ROM. Our low-rent approach to things will probably mirror the daily situations of other developers of this very populist medium.

Lastly, at the appropriate time, Whole Earth is in a prime position to spread the gospel of Wildcard/CD ROM to the network of its potential prime users. Our national, even global, audience will be Apple's. The wide circle of influential people we reach through the Catalogs, our Whole Earth Review magazine, and our teleconference network (The WELL) looks to us to point out the way. They trust our judgement about which new technologies are worthwhile. We would love to be able to say from personal experience that we think anyone contemplating CD ROM publishing should be using Apple's Wildcard/CD ROM set-up.

Timing

The $52,000 development cost reflects the demand of a short deadline. September is difficult but possible. October would be better. In either case, Whole Earth enjoys a reputation of always meeting its deadlines.

We anticipate enlarging the 1,000 card mini-Learning Disc into a mature, highly evolved Learning Disc that would demonstrate the malleable force of CD ROM. Loaded with 20,000 Wildcards it would make a considerable dent in our culture. If we conserved the momentum of the mini-disc construction and kept the rate of learning up, we could produce the 20,000-card version in an additional 6 months. Still figuring one hour per card, and our non profit rates, this would cost $700,000. We would consider that

amount to be an advance against royalties for the completed Apple/Whole Earth Learning Disc, version 1.0. The royalty is 15% of retail sales price, and would be counted against the units Apple gave away or bundled.

Whole Earth Explains Itself

For the past 18 years we have pushed the limits of paper technology to the extreme, distilling mammoth amounts of information into deep, easily accessible books. At its most robust, the Last Whole Earth Catalog grew to 614 oversize tabloid pages, weighing in at 5 1/2 pounds. It recommends the best tools and ideas in an innovative range from cybernetics to beekeeping. In all, its editions sold over 2 million copies.

The Catalog's success invented a new genre of publishing: the trade paperback. And a new medium: the nationally distributed small press. A Whole Earth /Wildcard CD ROM would again spearhead a wholly new medium.

The Catalog spawned (and encouraged) not only imitations of itself (The Whole Birth Catalog, The Whole Ocean Catalog, etc.), but gave life to thousands of small, special interest publishers who bypassed the production and distribution system of New York. The how-to books they developed met a hunger for vernacular instruction that Big Name publishers were not providing. They carried raw data, without gourmet packaging, in a format designed for exploration. The explosion of grassroots computer games (none coming from New York) is a witness to this continuing thirst.

We see our on-going job as trackers and filters of real news -- the kind of information that can change the user's behavior. The reader gets live news, attached to a live address. Each item inspires you to follow it up and try it out -- especially things you hadn't even imagined before. That's the sort of mentor you would love to find in an optical disc.

Even more than the paper Catalogs, the Learning Disc will cultivate participation. The hand, rather than just the eye, will choose. The user's intelligence steers by stretching an arm to slide the mouse, uncovering a connection. Brainstorming becomes a motor skill.

Practicing is a lot of what learning is about, in particular practicing in the unknown. Like the Catalogs, the Learning Disc exercises the unexpected. Again and again the steps for comprehending alien entities like mathematical statistics, or boat building, are mapped out

visually. When the student can come upon a new territory and navigate by rote, you've got a life-long learner.

The original Catalog was self-published using low-rent technology -- a Polaroid MP-3 camera and an IBM Selectric Composer. Current editions are still constructed with personal, flexible tools. The Whole Earth crew is at their best when inventing new medium channels as well as new content. Trailblazing through the bugs, quagmires and still oozing protocol conventions of the new CD ROM medium is the kind of pioneering we've done successfully before.